July 4, 1959

Dear Ellis:

Thank you for the ms. It is a very nice job of work, and we all enjoyed reading it. You did ask for comments-- I have just a couple. First of all, it would be a rounder job if the enzymology were in the same paper, provided it's ready. We have been waiting to write a detailed account of the Gal work, and I think we may have erred in the opposite direction. The second point is that you may have overlooked just these studies, which are quite concordant with your own. Unfortunately, Esther hasn't yet tied it all together except in the enclosed abstract; there are also some published remarks by Kalckar in the Symp. Chemical Basis of Heredity, and in a paper by Kurahashi in SCIENCE (Jan 18 '57). Esther wishes she could be so successful in linear mapping, but she has had too much trouble from coincidence; she's trying other approaches now. What do you make of your group C mutants? There is something possibly analogous in $Gal_3$ and $Gal_9$ which seem to be defective (acc. Kalckar) in all three enzymes.

We have been collecting some Ara mutants in K-12 with a somewhat similar program in mind, stressing mating analysis and interaction in heterozygotes, however. There are at least two cistrons between T and L (exactly agreeing with your findings; not improbably a third one). We have done nothing yet on physiological characterization, however, waiting to get settled here and for the arrival of Dick Soffer, a postdoct. fellow this fall. We will have to think about further strategy now, but I think it would be important to round out this kind of story with heterozygote analysis.

We have found one or two mutants in a completely different region (probably near S); however, these are not complete Ara-negatives and may prove to be permease effects. I suppose I should wait for your second paper, but I should ask about the inducibility of your enzymes. Group C might prove to be a permease effect, the kinase and [illegible]erase then being deficient owing to a failure in sequential induction. From the presence of the [illegible] kinase in group A mutants, we might have to infer that L-arabinose rather than L-ribulose induces the kinase. You can check some of these points nicely by feeding L-ribose or L-ribulose (if you have enough of these compounds!) Have you found any defects for the 4-epimerase step from L-ribulose-5-phosphate to D -xylulose-5-phosphate?

A propos some details on the mapping: I find it very easy to confuse myself while handling such data, but shouldn't the headings for the last three columns of table 1 read $leu^+$, $thr^+$ and $ara^+$ respectively? (Perhaps one should read in the implication unselected marker from the donor-- but this is hard enough without allowing any ambiguity I am indicating on the enclosed thermofax how one might write such tables in a notation I hope to propose. One goes through this process mentally anyhow whenever one looks at linkage data, so why not write it down. It's simple enough.

Table 2 would be clarified by repeating the statement (given in text) that all the tests are thr+ Leu+ --x thr- leu-. I assume that the + superscripts have also been ommitted in table 3. I did not give this the 'full treatment' along the lines of table 2, but recommend this to you. Also I feel quite strongly that you should present all of your data, amplifying table 4. The suggested notation may help the presentation. Other workers may find implications in such data that you don't see yourself-- I know that Cavalli, for example, would very much like to have full presentations to try some quantitative biometrical theory on. What you are calling negative interference (or coincidence) may be quite important in analysing the possible distribution in the size of the exogenotes, and this is at least one item that might be better gleaned from an exhaustive tabulation. This table might be simplified by putting

reciprocal transductions on the same line of the table, and always writing the standard test in the order you have already inferred for the markers.(which I see you have done in effect).
E.G. table 3 might begin to look like: (You can improve the wording of the headings).

| Cross | Percent leu+ ara+ /ara+ | Number ara+ tested | Reciprocal percent | number |
|---|---|---|---|---|
| 2 x- 13 | 24.2 | 62 | 47.7 | 218 |
| 13 x- 7 | 31.9 | 216 | 56.8 | 199 |

This eliminates some of the redundant numbers; table 3 as you present it could be reconstructed from this form which is a little more compact. If you want to go one step further you could again calculate leu+ ara+ / leu- ara- , i.e., for 2 x- 13 this would be 24.2/75.8 = .319 . All the information is still there. I don't see much point in saying how many experiments were pooled (viz. your parenthes is (1) or (2)) unless you give the numbers from each separately, as you have done in table 5,commendably.

I must admit I haven't gone over the entire set of data to verify your order-- one reason is that I hoped I might have persuaded to present them in this form, so that your inferences would be even more self-evident. As I've already emphaiszed the notation proposed just follows through what one has to do mentally anyhow, and to that extent should help.

I am enclosing a summary of Demerec's opus magnus-- the table has in it <u>all</u> the data which are presented in 36 of his tables. A point to stress is that one can now more easily see some weaknesses in his argument; e.g., it looks as if the sequence of steps ABCD depends crucially on experiment 10, and on the difference between 23 vs. 26 in reciprocal transductions!

The 1111, 0000 notation makes obvious what I went to too much trouble at in a note in MGB. In qualitative transduction mapping, the main point is that the 101 type of recombinant should be forbidden. The 010 type is often relatively frequent, and I think this may argue for a distribution of exogenote sizes, some of them being --1-- to begin with. One can also argue that the full genotype is better written ...0000000<u>101</u>00000000 versus ....0000000<u>10</u>0000000000, so that the difference reflects the incidence of 4 exchanges ..01010... versus the minimum of 2 ...00100. (This on the model of transduction where the exogenotes are uniform --<u>111</u>-- -x 0000<u>0000</u>000000).

I am planning to this up very soon for American Naturalist under the title "Notation for Genetic recombination analysis." It has been lying fallow for four years now, which is long enough. I did use it in PNAS Dec. 57.

Best wishes,

Joshua Lederberg